Filter of noisy annotation for Semantic Segmentation in Satellite Images

Wun-Cin Lian(練文勤), Li-Wei Chen(陳立緯). Department of Mathematics, National Chiao-Tung Univ. , Taiwan (ctw@math.nctu.edu.tw)
Advisor: *Chin-Tien Wu(吳金典),*

Abstract

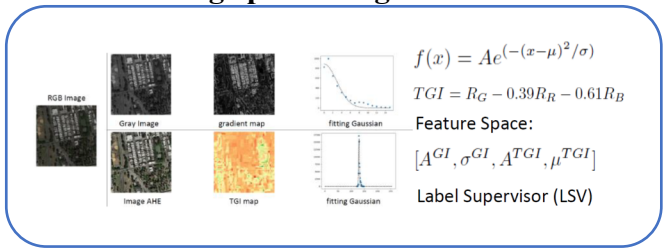
In semantic segmentation of satellite images, the task to acquire accurate label is tedious and time consuming. Automation of such tasks turns out to be challenging due to the fact that existing labels is generally quite coarse and noisy. Weakly-supervised learning recently proposed by Adrien and Hicham et al [1] shows some improvement on accuracy but is still unsatisfactory in practice. In this work, we propose to constrain the segmentation network such as U-Net [2] on multi-level convolution neural networks (CNNs) for binary urban semantic segmentation in satellite images to filter noisy labels.

Problem description

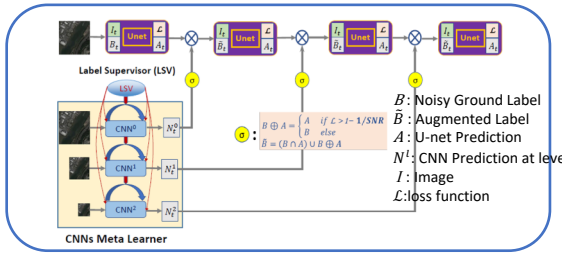
To deal with noisy annotations, we filter noisy data by the proposed label supervisor. Then, the prediction of CNN which trained by filtered data constrains U-net to improve performance.

Method

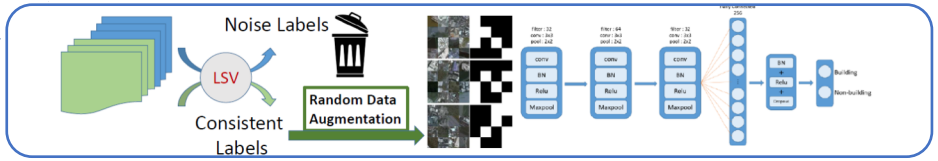
Data and Image processing:



CNNs constrained U-net:



Label filtering for CNNs training:



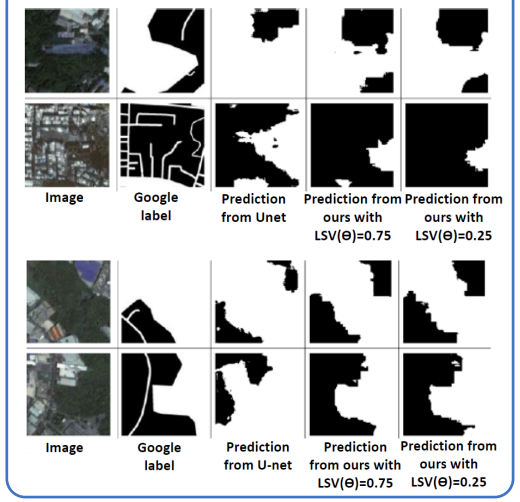
Discussion

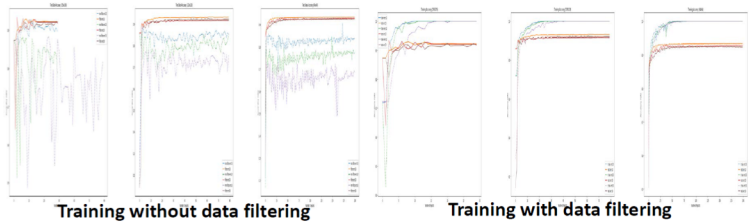
The proposed label supervisors, which learn from image features, efficiently filter the annotation noise. With these supervisors, the accuracy of trained model with 30% noise is significantly increased by at most 24% (from 68.75% to 92.33%), compared to that without the label supervisors.

Conclusions

In this work, we define the image-level label from pixel-level label if the image annotation has more pixels with building label. Otherwise, the image is labeled as non-building. Since there is no guarantee one time-consistency between the image and the ground truth annotation, to test our algorithm, we add 10%, 20%, 30% more noisy image-level labels to the dataset. Our results show that the CNNs-constrained U-net is noise-tolerant with accuracy more than 90% on image-level label predictions. Our method improves semantic segmentation from the U-net trained with noisy labels in our test images. Furthermore, accurate urban region masks can be created for images taken from Taiwan's Formosa V satellite.

Results

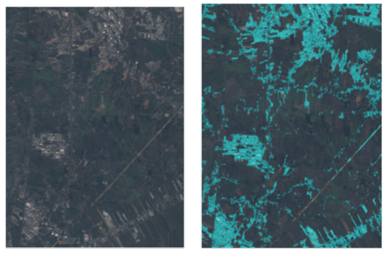




Accuracy	Before Filtering		After Filtering	
Noise	20%	30%	20%	30%
CNN0	78.74%	81.98%	92.09%	91.80%
CNN1	85.13%	75.91%	91.63%	92.25%
CNN2	77.26%	68.75%	92.81%	92.33%

Accuracy Comparison

Dataset:
Collected by Phillip et al [3] in their pix2pix project. The dataset is a selection from google API which contains 20000 RGB satellite images, with 5cm GSD resolution and size 512x512. In this project, the size is cropped as 256x256.



Reference

[1] Weakly Supervised Semantic Segmentation of Satellite Images, Adrien and Hicham , JURSE (2019)

[2] U-net: Convolutional networks for biomedical image segmentation, Ronneberger et al. MICCAI (2015)

[3] Image-to-image translation with conditional adversarial networks, Phillip et al. CVPR (2017)